BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

COVID-19 outbreak rates and infection attack rates associated with the workplace: a descriptive epidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055643
Article Type:	Original research
Date Submitted by the Author:	19-Jul-2021
Complete List of Authors:	Chen, Yiqun; Health and Safety Executive Aldridge, Timothy; Health and Safety Executive Ferraro, Claire; Public Health England South Region Khaw, Fu-Meng; Public Health Wales
Keywords:	Epidemiology < INFECTIOUS DISEASES, COVID-19, Public health < INFECTIOUS DISEASES, OCCUPATIONAL & INDUSTRIAL MEDICINE

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

COVID-19 outbreak rates and infection attack rates associated with the workplace: a descriptive epidemiological study

Yiqun Chen[1], Timothy Aldridge[1], Claire F. Ferraro[2], Fu-Meng Khaw[3]

1 Health and Safety Executive, Buxton, SK17 9JN, UK
2 Public Health England South West, Public Health England, Bristol, BS1 6EH, UK
3 Health Protection and Screening Services, Public Health Wales, Cardiff, CF10 4BZ, UK

Correspondence to Dr Yiqun Chen, Health and Safety Executive, Buxton, SK17 9JN, UK;
yiqun.chen@hse.gov.uk

ABSTRACT

Objectives

A large number of COVID-19 outbreaks/clusters have been reported in a variety of workplace settings since the start of the pandemic but assessment of the rate of outbreak occurrence in different types of workplace settings has not previously been assessed. The objectives of this paper are to identify the geographical areas and industrial sectors with a high rate of outbreaks of COVID-19; and compare infection attack rates by enterprise size and sector in England.

Methods

Public Health England (PHE) HPZone data on COVID-19 outbreaks in workplaces, between 18 May – 12 October 2020, were analysed. The workplace outbreak rates by region and sector were calculated, using National Population Database (NPD) on the total number of workplaces as the denominator. The infection attack rates were calculated by enterprise size and sector, using PHE Situations of Interest (SOI) data on the number of test-confirmed COVID-19 cases in a workplace outbreak as the numerator, and using NPD data on the number employed in that workplace as the denominator.

Results

The average outbreak rate was 66 per 100,000 workplaces. Of the nine English regions, the North West had the highest workplace outbreak rate (155/100,000 workplaces). Of the industrial sectors, manufacturers and packers of food had the highest outbreak rate (1,672/100,000), which was consistent across seven of the regions. In addition, high outbreak rates in warehouses were observed in the East Midlands and the North West. The overall median attack rate was 3.4%. The highest attack rate was for outbreaks in close contact services (median 16.5%), followed by outbreaks in restaurants and catering (median 10.2%), and in manufacturers and packers of non-food products (median 6.7%).

Conclusions

Early identification of geographical regions and industrial sectors with higher rates of COVID-19 workplace outbreaks can inform interventions to limit transmission of SARS-CoV-2.

For peer review only

Strengths and limitations of this study

- A large number of workplace outbreaks recorded in the national public health information system were analysed and the findings can guide interventions to target high risk areas and industrial sectors and to limit the spread of the SARS-CoV-2 virus.
- High rates of outbreaks were found in sectors where production demands are high, and workers cannot work from home, which indicate the importance of monitoring outbreak rates by industrial sector to identify emerging high risk areas as the country is moving out of the pandemic and more sectors are increasing their work capacity.
- The same methodological approach can be applied to the calculation of outbreak rates and attack rates in other geographical locations and other types of settings to support pandemic response.
- The number of outbreaks included in the analysis could be affected by national and local level operational changes and biased towards large and more impactful outbreaks. These will under-estimate the true outbreak rates and limit the ability to measure regional variations.
- The working population would be overestimated in some workplace settings with reduced number of employees working during the pandemic but underestimated in other workplace settings with many seasonal workers, agency workers and sub-contractors. These will cause imprecisions in the attack rate calculation.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly transmissible novel virus that has caused a pandemic of the "coronavirus disease 2019" (COVID-19)[1]. On 30th January 2020, the World Health Organisation declared COVID-19 as a public health emergency of international concern, and later declared a pandemic on 11th March 2020[2]. COVID-19 is a highly contagious disease and can spread rapidly without effective control measures. Because of the heterogeneity characteristics of the SARS-CoV-2 transmission, COVID-19 cases are appearing in clusters in different settings [3 4].

In October 2020, Public Health England reported 503 COVID-19 outbreaks/clusters in workplace settings in the previous 4 weeks. This is compared to 720 in care homes, 853 in education settings and 89 in hospital settings[5]. A survey conducted by the European Centre for Disease Prevention and Control (ECDC) reported a total of 1377 COVID-19 clusters in workplace settings across 13 EU/EEA countries and the UK between March and July 2020. Most clusters were reported in long-term care (591 clusters) and hospital (241 clusters) settings, followed by food packing and processing (153 clusters), non-food manufacturing (77 clusters) and office settings (65 clusters)[6]. However, the total number of settings and the number of people exposed within these settings (i.e. the denominator data) could vary significantly. Without the denominator information to calculate the rate of outbreaks, it is difficult to know which types of settings are more likely to experience an outbreak.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

This study aims to analyse the occurrence of COVID-19 outbreaks in workplace settings in England to understand which industrial sectors are more likely to experience an outbreak and to estimate the potential extent of the transmission in these workplace outbreaks. These will guide further research and control measures. However, the design of this study would not allow the investigation of factors potentially contributing to the outbreaks. A separate study is underway to address this[7].

This study is part of the United Kingdom National Core Study on Transmission and Environment[8]. The Health and Safety Executive (HSE) and Public Health England (PHE) worked collaboratively and, with the appropriate data-sharing agreements in place, linked the relevant datasets to calculate the outbreak rates for different workplace settings and the infection attack rates among workers working in these outbreak settings.

METHODS

Public Health England (PHE) data on COVID-19 outbreaks in the workplace, between 18 May and 12 October 2020, were analysed. The workplace settings here are defined using the categories in PHE's surveillance system. They include non-residential settings that are not schools or hospitals, as outbreaks in these settings are recorded and analysed separately[9].

A COVID-19 cluster is defined as two or more test-confirmed cases of COVID-19 among individuals associated with a setting (i.e. a workplace) with onset dates within 14 days, where information about exposure between the confirmed cases is not available or is absent. A COVID-19 outbreak is a COVID-19 cluster where direct exposure between at least two of the test-confirmed cases can be identified or information on an alternative source of infection outside the setting is absent for the initial identified cases[10].

Data from three sources, namely PHE HPZone dataset, PHE Situations of Interest (SOI) dataset and the HSE National Population Database (NPD), were used to calculate: 1) outbreak rates by geographical area (Regional and Upper Tier Local Authority (UTLA) and industrial sector; and, 2) attack rates of individual workplace outbreaks by enterprise size (small, 1-49 employees; medium, 50-249 employees; and large, 250 employees or more) and industrial sector. These three data sources are described in more detail in below.

HPZone dataset

HPZone is a national web-based system for communicable disease control in England and is PHE's case management system[11]. It has direct import of laboratory data, receiving statutory infectious disease notifications and collecting contextual data of management of infectious disease cases and outbreaks, and other non-infectious environmental threats. During the COVID-19 pandemic, HPZone provides summary-level information about the COVID-19 situations (i.e. outbreaks/clusters) that local Health Protection Teams (HPTs) are responding to. HPTs receive information about suspected or confirmed cases of COVID-19 directly from workplaces or through 'coincidence reports' from NHS Test and Trace, where two or more individuals report in the same workplace. Test-confirmed cases are linked to HPZone through the Second-Generation Surveillance System (SGSS), which is the national laboratory reporting system used in England to capture routine laboratory data, including data on infectious diseases. The HPZone data are verified by epidemiologists from the PHE National

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Surveillance Cell if a situation is a confirmed outbreak or a cluster of COVID-19. This is done at a snapshot in time on a weekly basis for the previous week's new situations. Outbreaks evolve over time. If the information about these outbreaks is not updated, for example to capture the increased number of confirmed cases as the outbreak develops, the data could underestimate the true size of the outbreak or clusters as more data become available over time about these outbreaks.

Situations of Interest (SOI) dataset

The Situations of Interest (SOI) dataset is a subset of outbreaks from the HPZone dataset that are deemed to be more complex to manage and includes updates on the number of test-confirmed COVID-19 cases as the outbreaks evolve over time. At the time of the data analysis, there was no formal definition of a SOI. It is used operationally to share understanding of significant outbreaks due to their scale, impact and complexity. A SOI outbreak will be updated regularly until transmission is controlled and as such provides a dynamic tool to track the total number of confirmed cases for the outbreak.

National Population Database (NPD)

The National Population Database (NPD) includes geographically referenced estimates of the Great Britain (GB) population in Geographic Information System (GIS) layers [12]. The NPD groups the GB population into five themes: Residential, Sensitive (e.g. schools, care homes, hospitals and prisons), Transport, Workplaces and Leisure. The Workplace layer provides information on individual workplaces including the number of employees, industry type (using Standard Industrial Classifications (SIC)) and a spatial reference (address and postcode). The workplace information is extracted from the Office for National Statistics (ONS) Inter-Departmental Business Register (IDBR)[13] at enterprise level, with the data used in this analysis extracted in May 2019. This extract included information for 2 million UK businesses.

Outbreak rates and attack rates

Outbreak Rate = the number of outbreaks in workplace settings / 100,000 workplaces

Outbreak rate is defined as the proportion of workplace settings with COVID-19 outbreaks, expressed as the number of outbreaks per 100,000 workplaces. The numerator is the number of confirmed workplace outbreaks identified from HPZone. The denominator is the total number of workplaces identified from the NPD.

Attack Rate = the number of test-confirmed COVID-19 cases in a workplace outbreak setting / 100 employed in that setting

An attack rate measures the proportion of persons in an identified population who become infected during an outbreak[14]. It indicates the potential extent of the transmission in an outbreak. It is defined here as the proportion of workers in a workplace that become cases of COVID-19 by the end of the outbreak, expressed as a percentage. The numerator is the number of test-confirmed COVID-19 cases in a workplace outbreak obtained from the SOI dataset. The denominator is the number employed in that workplace obtained from the NPD.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The list of outbreaks/clusters in the HPZone dataset and the SOI dataset are categorised into primary, secondary and tertiary contexts. Workplace is one of the primary contexts, for which the secondary contexts (categories) and the tertiary contexts (sub-categories) are listed in Table 1. All secondary contexts were included as sectors and were mapped against the Standardised Industry Classifications (SIC) before matching them to the denominator dataset.

Table 1 Public Health England (PHE) classification of workplace settings, July 2020

Workplace setting category/Sector	Sub-category
Primary producers	Fruit and vegetable growers, animal and animal producers
Manufacturers and packers of food	Abattoir, meat products, alcoholic beverage, non-alcoholic beverage, dairy produce, fruit and vegetables, bakery, confectionery, ready meals, and other
Manufacturers and packers of non-food	Textiles and garments, electronics, car manufacturer, furniture, chemical plant, pharmaceuticals, printing, and engineering
Warehouses	
Distributors and transporters	Wholesalers, haulage company, and food distributors
Retailers	Supermarket, small retailers and other
First responders	Ambulance, fire services, police
Military sites	Army, navy and air force
Restaurants and caterers	Restaurant/café/canteen, hotel/guest house, pub/club, take-away, mobile food unit, and other
Offices	
Close contact services	Hairdressers, barber shops, beauty and nail bars, make-up studios, tattoo studios, tanning salons or booths, spas and wellness businesses, sports and massage therapy, wellbeing and holistic locations, dress fitters, tailors, and fashion designers
Other	

Outbreak sites from the SOI records were linked to workplaces in the NPD through their postcode and business name. Unmatched SOI records were not included in the attack rate analysis. Furthermore, if the number of cases exceeded the number employed, the sites were excluded from the analysis. This may be due to underestimation of employment in the NPD for some workplace settings, such as crop production and warehouses where there is a reliance on temporary agency worker. Geographical coordinates were added to HPZone and SOI data from the ONS Postcode Directory[15], using the postcode of the outbreak settings. Statistical software R was used for the analysis and record linkages; Microsoft Excel was used for data preparation and creating charts. ArcGIS was used to create maps.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Patient and public involvement

Patients and the public were not involved in the design or conduct of the study.

RESULTS

In total, 1,317 confirmed workplace outbreaks were identified from HPZone, of which 1,305 could be mapped to NPD by postcode. In addition, 390 outbreaks were identified from the SOI dataset, of which 285 could be linked directly to records in the NPD workplaces to add SIC and employment information. A further 21 outbreaks from the SOI dataset, where no case numbers were recorded or where the number of cases exceeded the number employed, were removed. This leaves 264 SOI records of outbreaks, including a total 2,649 confirmed COVID-19 cases, for the attack rate calculation. See online supplemental figure S1 on the geographical distribution of the outbreaks.

Outbreak rates by geographical area (Region, UTLA)

Of the nine regions in England, the North West had the highest number of outbreaks, affecting 351 workplaces, as well as the highest rate of outbreaks (155/100,000 workplaces) (Table 2)**.** Of the 151 Upper Tier Local Authorities (UTLAs), the largest numbers of workplace outbreaks were mainly observed in northern English towns and cities with the highest outbreak rates registered in Blackburn with Darwen (387/100,000), Sandwell (351/100,000), Liverpool (349/100,000), Rochdale (277/100,000), Manchester (275/100,000) and Bradford (254/100,000).

Table 2 Number and rate of COVID-19 workplace outbreaks by English Region, May-Oct 2020

Region	Number of Outbreaks	Number of Workplaces	Outbreak Rate (per 100,000)
North West	351	226,576	155
Yorkshire and The Humber	198	168,184	118
West Midlands	215	183,534	117
East Midlands	134	156,900	85
North East	39	67,056	58
London	149	375,249	40
South West	84	215,640	39
East of England	71	226,190	31
South East	64	349,945	18
Total	1,305	1,969,274	66

Outbreak rates by sector

In comparison with other sectors, retailers had the highest number of outbreaks, affecting 219 workplaces, followed by manufacturers and packers of non-food products (195) and offices (193). However, after applying the denominator data, the highest outbreak rate was in manufacturers and packers of food (1,672/100,000), based on 117 outbreaks out of 6,998 workplaces. This was much higher than the outbreak rates for the remaining sectors with warehouses and manufacturers and packers of non-food products the next highest at 385 per 100,000 workplaces and 308 per 100,000 workplaces respectively (Table 3).

Table 3 Number and rate of workplace outbreaks by sector in England, May-Oct 2020

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Sector	Number of Outbreaks	Number of Workplaces	Outbreak Rate (per 100,000)
Manufacturers and packers of food	117	6,998	1,672
Warehouses	58	15,058	385
Manufacturers and packers of non-food	195	63,312	308
Retailers	219	195,025	112
First responders/Military sites	57	67,257	85
Distributors and transporters	84	125,414	67
Restaurants and caterers	53	117,836	45
Offices	193	721,351	27
Close contact services	13	52,866	25
No setting type assigned	54	511,071	11
Primary producers	8	93,086	9
Other	266	-	-
Total	1,317	1,969,274	67

Outbreak rates by region and sector

High outbreak rates in manufacturers and packers of food were observed consistently across seven regions, including: the West Midlands (3,555/100,000 workplaces), Yorkshire and the Humber (3,132/100,000 workplaces), the North West (2,926/100,000 workplaces), the East Midlands (2,031/100,000 workplace), the East of England (1,664/100,000), the North East (1,282/100,000 workplaces), and the South West (638/100,000 workplaces) (Table 4). In addition, high rates of outbreaks were observed in warehouse settings in the East Midlands and the North West with an outbreak rate of 1,524 per 100,000 workplaces and 793 per 100,000 workplaces respectively (Table 4).

Table 4 The top 10 outbreak rates by English region and sector combined, May-Oct 2020

Region - Sector	Number of Outbreaks	Number of Workplaces	Outbreak Rate (per 100,000)
West Midlands - Manufacturers and packers of food	23	647	3,555
Yorkshire and The Humber - Manufacturers and packers of food	28	894	3,132
North West - Manufacturers and packers of food	28	957	2,926
East Midlands - Manufacturers and packers of food	13	640	2,031
East of England - Manufacturers and packers of food	12	721	1,664
East Midlands - Warehouses	19	1,247	1,524
North East - Manufacturers and packers of food	4	312	1,282
North West -	65	8,074	805

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Manufacturers and packers of non-food			
North West - Warehouses	15	1,891	793
South West - Manufacturer and packers of food	6	940	638
…			

Attack rates by enterprise size

A minority (29%) of the outbreaks recorded in SOI were in small enterprises (<50 employees) but the proportion of small enterprises was higher for close contact services (83%) and restaurants and caterers (56%). The overall median attack rate was 3.4% for outbreaks in all enterprises. The median attack rate was 1.1% for outbreaks in large enterprises (250 employees or more), 4.3% in medium size enterprises (50-249 employees), and 17.8% in small enterprises (1-49 employees). The attack rates increased as the number employed at a workplace decreased.

Attack rates by sector

Outbreaks in close contact services had the largest attack rate (median 16.5%), which was based on 22 test-confirmed cases at 6 outbreak sites (Table 5). The attack rates were also high for outbreaks in restaurants and caterers (median 10.3%), based on 49 test-confirmed cases at 14 sites; and in manufacturers and packers of non-food products (median 6.7%), which was based on 270 cases at 29 sites. Most of the outbreaks (162 of 264 outbreaks) had an attack rate less than 6%. However, in a small number of outbreaks (57 of 264) the attack rate was over 15% (see online supplemental figure S2).

Table 5 Median attack rates of workplace outbreaks by sector in England, May-Oct 2020

	Individual Outbreaks				Workplaces		Attack Rate	
Sector	Total cases	Total sites	Cases per site		Number employed at outbreak sites		Cases per 100 employed	
			Median	IQR[(a)]	Median	IQR[(a)]	Median	IQR[(a)]
Close contact services	22	6	3	2	16	10	16.5	6.7
Restaurants and caterers	49	14	4	2	38	94	10.3	14.2
Manufacturers and packers of non-food	270	29	8	7	122	269	6.7	11.1
No setting type assigned	99	15	4	6	56	112	5.4	9.9
Retailers	115	28	4	2	120	242	4.9	16.1
Offices	133	23	5	4	133	207	4.3	15.7
Manufacturers and packers of food	1,384	79	7	12	423	641	2.3	7.0
First Responders/ Military sites	44	15	3	2	113	422	2.1	4.0
Other	109	24	3	3	169	241	2.0	10.6
Warehouses	104	12	3	8	579	781	1.6	1.2

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Distributors and transporters	193	16	4	9	650	693	1.2	4.7
Primary producers	127	3	3	61	*	*	*	*
Total	2,649	264	4	6	176	473	3.4	11.3

(a): Interquartile range (IQR)

*: the number of outbreak sites is too small to calculate.

DISCUSSION

Our study has utilised the number of confirmed COVID-19 outbreaks recorded in PHE information system and combined them with relevant denominator data held by HSE to calculate outbreak rates and attack rates by sector and geographical area. A relatively large number of outbreaks were observed in some workplace settings, including retail, manufacturers and packers of non-food products and offices. After applying the denominator data of the total number of the relevant settings, manufacturers and packers of food had the highest outbreak rates and this was consistent across seven English regions. Manufacturers and packers of food are part of the national infrastructure and these workplaces were kept open throughout the pandemic even during the national lockdown. Outbreaks of COVID-19 in manufacturers and packers of food have been frequently reported in the literature and in the media in many counties[16]. However, only a few studies have investigated the potential transmission risk factors in this type of workplace settings [17]. High rates of outbreaks were found in sectors where production demands are high, and workers cannot work from home. It will be important to continue to monitor outbreak rates by industrial sector as the country is moving out of the pandemic and more sectors are increasing their work capacity.

Our study has also utilised data from the public health COVID-19 outbreak management records to calculate infection attack rates. This allows comparison of the potential extent of transmission between outbreaks in different workplace settings. Close contact services and restaurants/caterers had the highest attack rates which were mostly associated with outbreaks in small enterprises. Manufacturers and packers of non-food products also had relatively large attack rates but were mostly associated with outbreaks in medium and large enterprises. However, it is worth noting that the SOI data are skewed towards large and more impactful outbreaks. Furthermore, more detailed analysis of attack rates is limited by low numbers of outbreaks in certain industrial sectors, such as primary producers which include fruit and vegetable growers, animal and animal products.

Our analysis carried some limitations. The potential under-identification of outbreaks in small enterprises (<50 employees) in the numerator coupled with the vast number of small enterprises in the denominator may greatly underestimate the outbreak rates. This could particularly impact on small business-dominated sectors, such as close contact services and restaurants/caterers, where estimated outbreak rates were relatively low, but attack rates were relatively high.

The number of outbreaks reported to HPZone could be affected by national and local level operational changes. For example, as caseload increased in September and October 2020, some HPTs transferred the management of some outbreaks/clusters to local authorities. As a result, HPZone no longer represents a comprehensive list of COVID-19 outbreaks/clusters in England. This

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

will affect the ability to measure the changes of outbreak occurrence or outbreak rate over time, as well as the ability to measure regional variations but it remains valuable to conduct sector comparisons.

SOI outbreak data is a subset of the HPZone outbreaks/clusters. Data entry was through a separate mechanism. The proportion of HPZone outbreaks/clusters in the workplace being reported as SOI decreased over time, especially from September 2020 onward as HPTs were under pressure to respond to an increasing number of outbreaks. However, it is unclear if these decreased are biased toward certain sectors. It remains valuable to assess the attack rates of individual outbreaks across different sectors.

NPD workplaces information also has some limitations in providing reliable working population data as the denominator, which will cause imprecisions in the attack rate calculation. In our study, NPD data represent the distribution of the GB population pre-pandemic; the number of employees in some workplace settings will be reduced during the pandemic due to social distancing measures. This may cause underestimation of the attack rates due to overestimation of the denominator. The level of underestimation varies by sector with some sectors completely closed and others kept operating in full capacity throughout the pandemic. However, the impact of this limitation may attenuate as society gradually opens.

In addition, the NPD workplaces information may not capture the number of employees in the transient workforce or working in irregular patterns, for example, seasonal workers in the agriculture sector. Employees in some other workplaces, such as in distribution centres, transportation of goods between depots, and in construction will be accounted for but their non-fixed working locations will not be well-represented by a single geographical reference (e.g. postcode of the company address). Similarly, agency staff and sub-contractors are unlikely to be accounted for at the location where they carry out their work activities. This may cause over estimation of the attack rate due to the underestimation of the denominator.

Early identification of COVID-19 outbreaks/clusters and visualisation of their geographical distribution can provide a rapid assessment of where the SARS-CoV-2 transmission is occurring. A large number of COVID-19 outbreaks/clusters have been reported, both in scientific literature and in the media, in a wide range of mostly indoor settings across the world [3 18]. Most of the COVID-19 clusters will be in residential settings, particularly in households, due to the increased risk of transmission caused by close and frequent contact[19]. However, a household cluster will not result in a large outbreak without the virus spreading beyond the household setting. Some of these individuals in households could also travel to other settings including the workplaces. Transmission is a continuous risk. It is difficult to establish where transmission really occurred. Community transmission will also occur through social gathering, particularly gathering outdoors, shopping in supermarkets or using public transport. However, it is difficult to identify outbreaks/clusters from the large number of transient populations in these settings without a rigorous surveillance system for widespread testing and detailed contract tracing. This may underestimate the relative importance of the potential transmission in these less well-defined settings or population.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Since our study, the approach of utilising the suitable denominator data to calculate outbreak rates has been adopted by the UK Joint Bio-security Centre (JBC) and will be embedded in their regular national surveillance analysis and reporting on workplace outbreaks and outbreaks rates. Although this study was only able to analyse the workplace outbreak data, the same approach can be applied to the calculation of outbreak rates and attack rates in other geographical locations and other types of settings, such as care homes, hospitals, schools and prisons. These will potentially guide interventions to target high risk areas and to limit the spreading of the virus.

This study was not able to assess the potential changes in COVID-19 outbreak rates and attack rates over time due to, in part, the limited time period of data and the inconsistency in recording outbreaks/clusters in the HPZone and SOI datasets. Further consideration will be to analyse the more enhanced outbreak/cluster data collected over time from NHS Test and Trace to identify past and emerging trends.

Evidence shows that there could be marked heterogeneity in the characteristics of SARS-CoV-2 transmission[4], with the majority (~80%) of the secondary transmission caused by a very small proportion of SARS-CoV-2 infected persons, and outbreaks of COVID-19 distributed unevenly in certain settings and geographical locations[20]. Our study has found increased rates of outbreak in certain industrial sectors and geographical regions, and a large variation of the size of the attack rates. The variation of the rates may be impacted by the type of work activities, the size of the enterprises, the transmission risk and the intervention strategies to limit the transmission in these sectors. The risk of transmission will also be associated with the behavioural and social factors of the individuals, the environment and the control measures that influence transmission dynamics of the virus in certain settings[3].

The current study has investigated the patterns and rates of COVID-19 outbreaks in England. Further studies, as part of the National Core Study programme, will investigate and identify the characteristics of the outbreak settings that could increase risk of transmission. A comprehensive epidemiological field study has been designed and commissioned to collect data from live COVID-19 outbreaks in workplace settings to better understand the transmission risk factors and transmission routes[8].

Contributors Conceptualization, Y.C., T.A., and F-M.K.; Methodology, T.A., Y.C., and C.F.F.; Validation, Y.C., T.A., C.F.F., and F-M.K.; Formal Analysis, T.A.; Investigation, Y.C., T.A., and C.F.F.; Data Curation, T.A. and C.F.F.; Writing-Original Draft Preparation, Y.C., T.A., and C.F.F.; Writing-Review & Editing, Y.C., T.A., C.F.F., and F-M.K.; Visualization, T.A.; Supervision, Y.C. and F-M.K.; Project Administration, F-M.K. and Y.C.; Funding Acquisition, UK COVID-19 National Core Studies Consortium.

Ethics approval This study constituted public health surveillance rather than research in human beings. Ethical approval from institutional review boards was not required. All information on individual persons had been anonymised.

Funding This work is supported by the UK Government COVID-19 National Core Study on Transmission and Environment

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Competing interests None declared.

Disclaimer The opinions and assertions contained herein are private views of the authors and do not necessarily reflect those of the Health and Safety Executive or Public Health England.

Copyright © Crown copyright (2021), Health and Safety Executive. This is an open access article distributed under the terms of the Open Government Licence v3.0, which permits re-use, distribution, reproduction and adaptation, provided the original work is properly cited.

Patient consent for publication Not required.

Data Availability Statement: The data used to support the findings of the study are included in the references within this paper. No additional data available.

For peer review only

REFERENCES

1. Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* 2021;19(3):141-54. doi: 10.1038/s41579-020-00459-7
2. Coronavirus disease 2019 (COVID-19): situation report, 51. Geneva: World Health Organization, 2020.
3. Leclerc Q, Fuller N, Knight L, et al. What settings have been linked to SARS-CoV-2 transmission clusters? [version 2; peer review: 2 approved]. *Wellcome Open Research* 2020;5(83) doi: 10.12688/wellcomeopenres.15889.2
4. Wong NS, Lee SS, Kwan TH, et al. Settings of virus exposure and their implications in the propagation of transmission networks in a COVID-19 outbreak. *The Lancet Regional Health - Western Pacific* 2020;4:100052. doi: https://doi.org/10.1016/j.lanwpc.2020.100052
5. PHE. National COVID-19 surveillance report: 2 October 2020 (week 40): Pubic Health England (https://www.gov.uk/government/publications/national-covid-19-surveillance-reports), 2020.
6. ECDC. COVID-19 clusters and outbreaks in occupational settings in the EU/EEA and the UK. Stockholm: European Centre for Disease Prevention and Control (ECDC), 2020.
7. Curran A. COVID-19 Outbreak investigation to Understand Transmission (COVID-OUT). 2020. https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/the-covid-out-study-covid-19/.
8. Theme 1: Rapid investigation of outbreaks and evidence synthesis; COVID-19 National Core Study on Transmission and Environment 2021 [Available from: https://sites.manchester.ac.uk/covid19-national-project/rapid-investigation-of-outbreaks-and-evidence-synthesis/.
9. PHE. Weekly national Influenza and COVID-19 surveillance report, Week 6 report (up to week 5 data) in 2021: Public Health England (https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports), 2021.
10. PHE. COVID-19: epidemiological definitions of outbreaks and clusters in particular settings, Public Health England: Pubic Health England; [updated 7 August 2020. Available from: https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters/covid-19-epidemiological-definitions-of-outbreaks-and-clusters-in-particular-settings#fnref:1.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

11. Kara-Zaïtri C, Gelletlie R, Schweiger M. The development and deployment of a national web-based system for communicable disease control in England. *International Journal of Infectious Diseases* 2012;16:e133. doi: https://doi.org/10.1016/j.ijid.2012.05.301
12. Gorce J-P, Aldridge, T., Forder, K., Holmes, W., McManus. H., and Purdy, K. The National Population Database: overview and developments. Health and Safety Executive Research Report RR1130, 2018.
13. Inter-Departmental Business Register (IDBR) Office for National Statistics (ONS) (https://www.ons.gov.uk/aboutus/whatwedo/paidservices/interdepartmentalbusinessregisteridbr), 2021.
14. CDC. Principles of Epidemiology in Public Health Practice: Centers for Disease Control and Prevention, https://www.cdc.gov/csels/dsepd/ss1978/glossary.html, 2012.
15. ONS. ONS Postcode Directory (November 2020). Open Geography Portal: Office for National Statistics (https://geoportal.statistics.gov.uk/datasets/ons-postcode-directory-november-2020), 2020.
16. Waltenburg MA, Rose CE, Victoroff T, et al. Coronavirus Disease among Workers in Food Processing, Food Manufacturing, and Agriculture Workplaces. *Emerg Infect Dis* 2021;27(1):243-49. doi: 10.3201/eid2701.203821
17. Guenther TaC-S, Manja and Indenbirken, Daniela and Robitailles, Alexis and Tenhaken, Peter and Exner, Martin and Ottinger, Matthias and Fischer, Nicole and Grundhoff, Adam and Brinkmann, Melanie,. Investigation of a superspreading event preceding the largest meat processing plant-related SARS-Coronavirus 2 outbreak in Germany *SSRN* 2020 doi: http://dx.doi.org/10.2139/ssrn.3654517
18. Liu T, Gong D, Xiao J, et al. Cluster infections play important roles in the rapid evolution of COVID-19 transmission: A systematic review. *International Journal of Infectious Diseases* 2020;99:374-80. doi: https://doi.org/10.1016/j.ijid.2020.07.073
19. PHE. National flu and COVID-19 surveillance reports, Public Health England: Public Health England; [updated 25 March 2021. Available from: https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports.
20. Frieden T, Lee C. Identifying and Interrupting Superspreading Events—Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Disease journal* 2020;26(6):1059. doi: 10.3201/eid2606.200495

For Peer Review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure S1 Geographical distributions of COVID-19 outbreaks associated with the workplace in England, May – Oct 2020

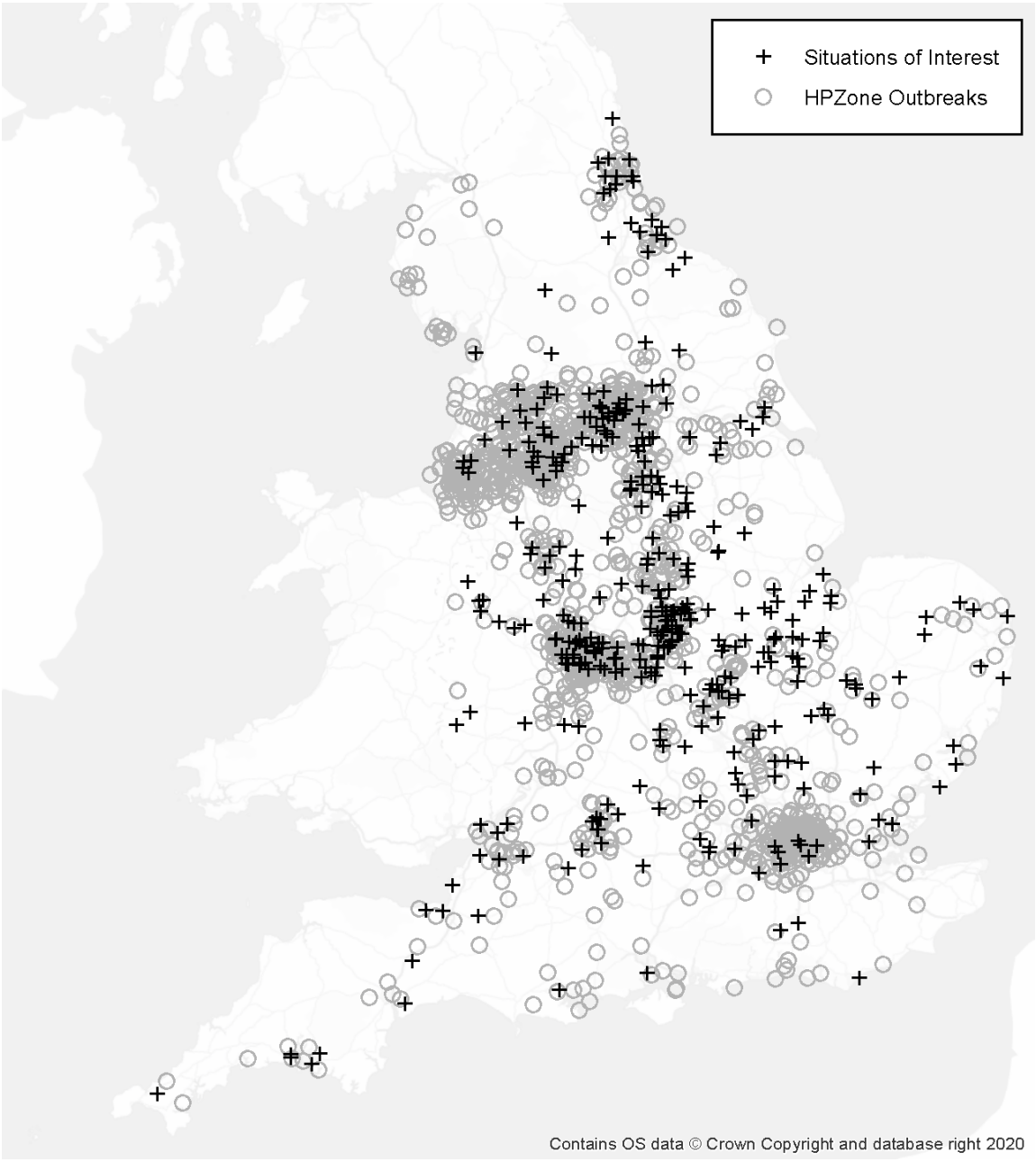
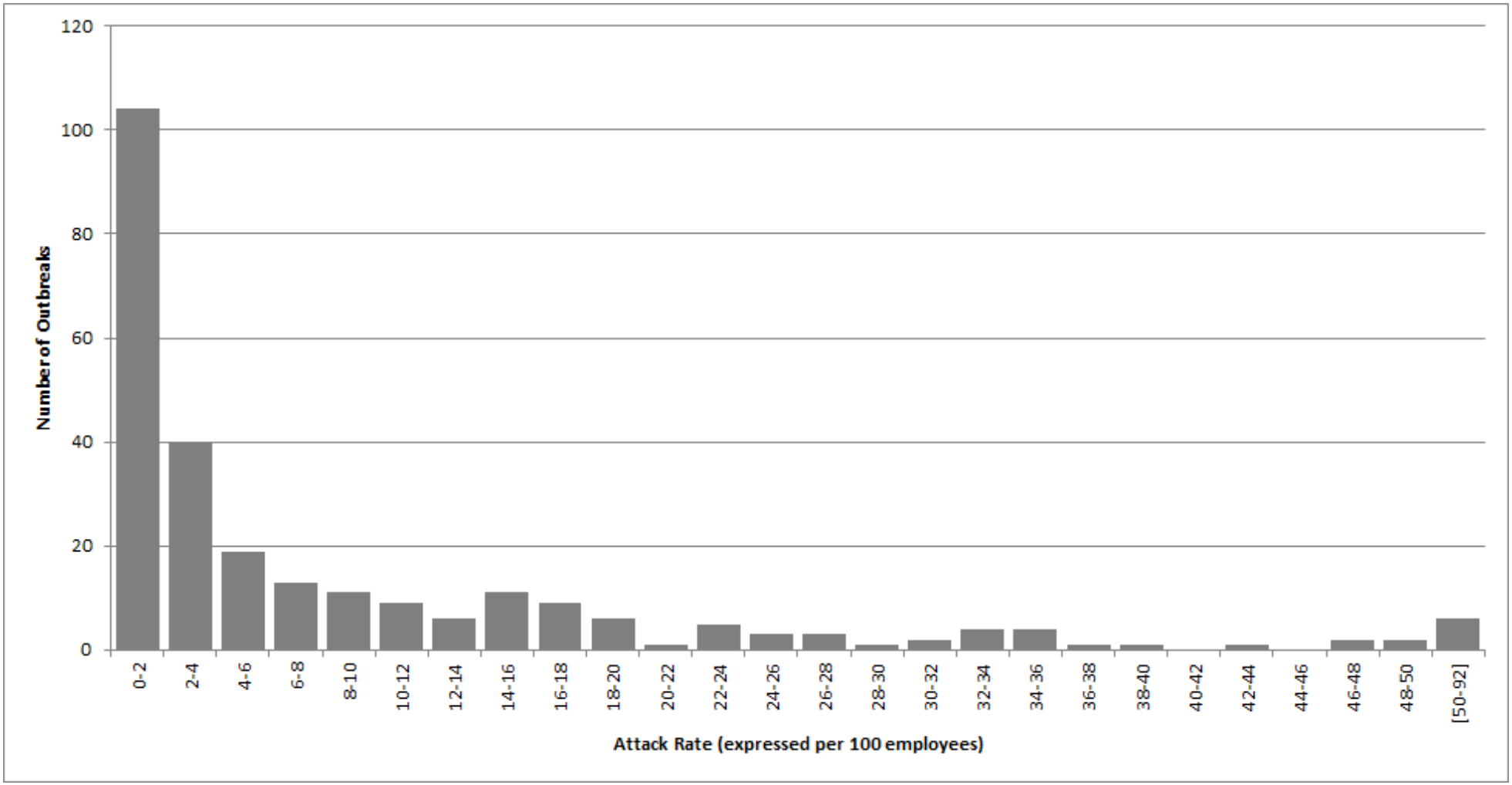


Figure S2 Distribution of the number of workplace outbreaks by attack rate in England, May-Oct 2020



The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 1-2	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Page 2
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 2-3		
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 3		
Methods					
Study Design	4	Present key elements of study design early in the paper	Page 3		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 3		

For peer review only

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	Page 3-5	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Page 3-5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Page 4-5	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Page 5
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 3-4		

For peer review only

Bias	9	Describe any efforts to address potential sources of bias	Page 5-6		
Study size	10	Explain how the study size was arrived at	Page 5-6		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Page 4-6		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Page 4-6		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	Page 3-5

For Peer review only

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Page 5-6
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Page 6	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Page 6
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	Page 6		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure	Page 6		

For peer review only

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Page 6-9		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			
Discussion					
Key results	18	Summarise key results with reference to study objectives	Page 9-10		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 10-11	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Page 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives,	Page 9-12		

For peer review only

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 11-12		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 12		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 12

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

*Checklist is protected under Creative Commons Attribution (CC BY) license.

For peer review only

BMJ Open

COVID-19 outbreak rates and infection attack rates associated with the workplace: a descriptive epidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055643.R1
Article Type:	Original research
Date Submitted by the Author:	04-Feb-2022
Complete List of Authors:	Chen, Yiqun; Health and Safety Executive Aldridge, Timothy; Health and Safety Executive Ferraro, Claire; Public Health England South Region Khaw, Fu-Meng; Public Health Wales
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Infectious diseases, Occupational and environmental medicine, Public health
Keywords:	Epidemiology < INFECTIOUS DISEASES, COVID-19, Public health < INFECTIOUS DISEASES, OCCUPATIONAL & INDUSTRIAL MEDICINE

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

COVID-19 outbreak rates and infection attack rates associated with the workplace: a descriptive epidemiological study

Yiqun Chen[1], Timothy Aldridge[1], Claire F. Ferraro[2], Fu-Meng Khaw[3]

1 Health and Safety Executive, Buxton, SK17 9JN, UK
2 Public Health England South West, Public Health England, Bristol, BS1 6EH, UK
3 Health Protection and Screening Services, Public Health Wales, Cardiff, CF10 4BZ, UK

Correspondence to Dr Yiqun Chen, Health and Safety Executive, Buxton, SK17 9JN, UK;
yiqun.chen@hse.gov.uk

ABSTRACT

Objectives

A large number of COVID-19 outbreaks/clusters have been reported in a variety of workplace settings since the start of the pandemic but the rate of outbreak occurrence in the workplace has not previously been assessed. The objectives of this paper are to identify the geographical areas and industrial sectors with a high rate of outbreaks of COVID-19; and compare infection attack rates by enterprise size and sector in England.

Methods

Public Health England (PHE) HPZone data on COVID-19 outbreaks in workplaces, between 18 May – 12 October 2020, were analysed. The workplace outbreak rates by region and sector were calculated, using National Population Database (NPD) on the total number of workplaces as the denominator. The infection attack rates were calculated by enterprise size and sector, using PHE Situations of Interest (SOI) data on the number of test-confirmed COVID-19 cases in a workplace outbreak as the numerator, and using NPD data on the number employed in that workplace as the denominator.

Results

The highest attack rate was for outbreaks in close contact services (median 16.5%), followed by outbreaks in restaurants and catering (median 10.2%), and in manufacturers and packers of non-food products (median 6.7%). The overall outbreak rate was 66 per 100,000 workplaces. Of the nine English regions, the North West had the highest workplace outbreak rate (155/100,000 workplaces). Of the industrial sectors, manufacturers and packers of food had the highest outbreak rate (1,672/100,000), which was consistent across seven of the regions. In addition, high outbreak rates in warehouses were observed in the East Midlands and the North West.

Conclusions

Early identification of geographical regions and industrial sectors with higher rates of COVID-19 workplace outbreaks can inform interventions to limit transmission of SARS-CoV-2.

Strengths and limitations of this study

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- This study has described in detail the relevant datasets used for the calculation of COVID-19 outbreak rates and infection attack rates in the workplace, in England, by industrial sector and geographical region.
- The same methodological approach can be applied to the calculation of outbreak rates and attack rates in other countries and for other types of settings to support pandemic response.
- The number of outbreaks included in the analysis was obtained from the public health outbreak management information system, which could be affected by national and local level operational changes and limit the ability to measure regional variations.
- The COVID-19 outbreaks included in the analysis could be biased towards large and more impactful outbreaks and, therefore, could under-estimate the true outbreak rates.
- The working population would be overestimated in some workplace settings with reduced number of employees working during the pandemic but underestimated in other workplace settings with many seasonal workers, agency workers and sub-contractors less likely to be accounted for, which could cause imprecisions in the attack rate calculation.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly transmissible novel virus that has caused a pandemic of the "coronavirus disease 2019" (COVID-19)[1]. On 30th January 2020, the World Health Organisation declared COVID-19 as a public health emergency of international concern, and later declared a pandemic on 11th March 2020[2]. COVID-19 is a highly contagious disease and can spread rapidly without effective control measures. Because of the heterogeneity characteristics of the SARS-CoV-2 transmission, COVID-19 cases are appearing in clusters in different settings [3 4].

In October 2020, Public Health England reported 503 COVID-19 outbreaks/clusters in workplace settings in the previous 4 weeks. This is compared to 720 in care homes, 853 in education settings and 89 in hospital settings[5]. A survey conducted by the European Centre for Disease Prevention and Control (ECDC) reported a total of 1377 COVID-19 clusters in workplace settings across 13 EU/EEA countries and the UK between March and July 2020. Most clusters were reported in long-term care (591 clusters) and hospital (241 clusters) settings, followed by food packing and processing (153 clusters), non-food manufacturing (77 clusters) and office settings (65 clusters)[6]. However, the total number of settings and the number of people exposed within these settings (i.e. the denominator data) could vary significantly. Without the denominator information to calculate the rate of outbreaks, it is difficult to know which types of settings are more likely to experience an outbreak.

This study aims to analyse the occurrence of COVID-19 outbreaks in workplace settings in England to understand which industrial sectors are more likely to experience an outbreak and to estimate the potential extent of the transmission in these workplace outbreaks. These will guide further research and control measures. However, the design of this study would not allow the investigation of factors potentially contributing to the outbreaks. A separate study is underway to address this[7].

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

This study is part of the United Kingdom National Core Study on Transmission and Environment[8]. The Health and Safety Executive (HSE) and Public Health England (PHE) worked collaboratively and, with the appropriate data-sharing agreements in place, linked the relevant datasets to calculate the outbreak rates for different workplace settings and the infection attack rates among workers working in these outbreak settings.

METHODS

Public Health England (PHE) data on COVID-19 outbreaks in the workplace, between 18 May and 12 October 2020, were analysed. The workplace settings here are defined using the categories in PHE's surveillance system. They include non-residential settings that are not schools or hospitals, as outbreaks in these settings are recorded and analysed separately[9].

A COVID-19 cluster is defined as two or more test-confirmed cases of COVID-19 among individuals associated with a setting (i.e. a workplace) with onset dates within 14 days, where information about exposure between the confirmed cases is not available or is absent. A COVID-19 outbreak is a COVID-19 cluster where direct exposure between at least two of the test-confirmed cases can be identified or information on an alternative source of infection outside the setting is absent for the initial identified cases[10].

Data from three sources, namely PHE HPZone dataset, PHE Situations of Interest (SOI) dataset and the HSE National Population Database (NPD), were used to calculate: 1) outbreak rates by geographical area (Regional and Upper Tier Local Authority (UTLA) and industrial sector; and, 2) attack rates of individual workplace outbreaks by enterprise size (small, 1-49 employees; medium, 50-249 employees; and large, 250 employees or more) and industrial sector. These three data sources are described in more detail in below.

HPZone dataset

HPZone is a national web-based system for communicable disease control in England and is PHE's case management system[11]. It has direct import of laboratory data, receiving statutory infectious disease notifications and collecting contextual data of management of infectious disease cases and outbreaks, and other non-infectious environmental threats. During the COVID-19 pandemic, HPZone provides summary-level information about the COVID-19 situations (i.e. outbreaks/clusters) that local Health Protection Teams (HPTs) are responding to. HPTs receive information about suspected or confirmed cases of COVID-19 directly from workplaces or through 'coincidence reports' from NHS Test and Trace, where two or more individuals report in the same workplace. Test-confirmed cases are linked to HPZone through the Second-Generation Surveillance System (SGSS), which is the national laboratory reporting system used in England to capture routine laboratory data, including data on infectious diseases. The HPZone data are verified by epidemiologists from the PHE National Surveillance Cell if a situation is a confirmed outbreak or a cluster of COVID-19. This is done at a snapshot in time on a weekly basis for the previous week's new situations. Outbreaks evolve over time. If the information about these outbreaks is not updated, for example to capture the increased number of confirmed cases as the outbreak develops, the data could underestimate the true size of the outbreak or clusters as more data become available over time about these outbreaks.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Situations of Interest (SOI) dataset

The Situations of Interest (SOI) dataset is a subset of outbreaks from the HPZone dataset that are deemed to be more complex to manage and includes updates on the number of test-confirmed COVID-19 cases as the outbreaks evolve over time. At the time of the data analysis, there was no formal definition of a SOI. It is used operationally to share understanding of significant outbreaks due to their scale, impact and complexity. A SOI outbreak will be updated regularly until transmission is controlled and as such provides a dynamic tool to track the total number of confirmed cases for the outbreak.

National Population Database (NPD)

The National Population Database (NPD) includes geographically referenced estimates of the Great Britain (GB) population in Geographic Information System (GIS) layers [12]. The NPD groups the GB population into five themes: Residential, Sensitive (e.g. schools, care homes, hospitals and prisons), Transport, Workplaces and Leisure. The Workplace layer provides information on individual workplaces including the number of employees, industry type (using Standard Industrial Classifications (SIC)) and a spatial reference (address and postcode). The workplace information is extracted from the Office for National Statistics (ONS) Inter-Departmental Business Register (IDBR)[13] at enterprise level, with the data used in this analysis extracted in May 2019. This extract included information for 2 million UK businesses.

Outbreak rates and attack rates

Outbreak Rate = the number of outbreaks in workplace settings / 100,000 workplaces

Outbreak rate is defined as the proportion of workplace settings with COVID-19 outbreaks, expressed as the number of outbreaks per 100,000 workplaces. The numerator is the number of confirmed workplace outbreaks identified from HPZone. The denominator is the total number of workplaces identified from the NPD.

Attack Rate = the number of test-confirmed COVID-19 cases in a workplace outbreak setting / 100 employed in that setting

An attack rate measures the proportion of persons in an identified population who become infected during an outbreak[14]. It indicates the potential extent of the transmission in an outbreak. It is defined here as the proportion of workers in a workplace that become cases of COVID-19 by the end of the outbreak, expressed as a percentage. The numerator is the number of test-confirmed COVID-19 cases in a workplace outbreak obtained from the SOI dataset. The denominator is the number employed in that workplace obtained from the NPD.

The list of outbreaks/clusters in the HPZone dataset and the SOI dataset are categorised into primary, secondary and tertiary contexts. Workplace is one of the primary contexts, for which the secondary contexts (categories) and the tertiary contexts (sub-categories) are listed in Table 1. All secondary contexts were included as sectors and were mapped against the Standardised Industry Classifications (SIC) before matching them to the denominator dataset.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 1 Public Health England (PHE) classification of workplace settings, July 2020

Workplace setting category/Sector	Sub-category
Primary producers	Fruit and vegetable growers, animal and animal producers
Manufacturers and packers of food	Abattoir, meat products, alcoholic beverage, non-alcoholic beverage, dairy produce, fruit and vegetables, bakery, confectionery, ready meals, and other
Manufacturers and packers of non-food	Textiles and garments, electronics, car manufacturer, furniture, chemical plant, pharmaceuticals, printing, and engineering
Warehouses	
Distributors and transporters	Wholesalers, haulage company, and food distributors
Retailers	Supermarket, small retailers and other
First responders	Ambulance, fire services, police
Military sites	Army, navy and air force
Restaurants and caterers	Restaurant/café/canteen, hotel/guest house, pub/club, take-away, mobile food unit, and other
Offices	
Close contact services	Hairdressers, barber shops, beauty and nail bars, make-up studios, tattoo studios, tanning salons or booths, spas and wellness businesses, sports and massage therapy, wellbeing and holistic locations, dress fitters, tailors, and fashion designers
Other	

Outbreak sites from the SOI records were linked to workplaces in the NPD through their postcode and business name. Unmatched SOI records were not included in the attack rate analysis. Furthermore, if the number of cases exceeded the number employed, the sites were excluded from the analysis. This may be due to underestimation of employment in the NPD for some workplace settings, such as crop production and warehouses where there is a reliance on temporary agency worker. Geographical coordinates were added to HPZone and SOI data from the ONS Postcode Directory[15], using the postcode of the outbreak settings. Statistical software R was used for the analysis and record linkages; Microsoft Excel was used for data preparation and creating charts. ArcGIS was used to create maps.

Patient and public involvement

Patients and the public were not involved in the design or conduct of the study.

RESULTS

In total, 1,317 confirmed workplace outbreaks were identified from HPZone, of which 1,305 could be mapped to NPD by postcode. In addition, 390 outbreaks were identified from the SOI dataset, of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

which 285 could be linked directly to records in the NPD workplaces to add SIC and employment information. A further 21 outbreaks from the SOI dataset, where no case numbers were recorded or where the number of cases exceeded the number employed, were removed. This leaves 264 SOI records of outbreaks, including a total 2,649 confirmed COVID-19 cases, for the attack rate calculation. See online supplemental figure S1 on the geographical distribution of the outbreaks.

Outbreak rates by geographical area (Region, UTLA)

Of the nine regions in England, the North West had the highest number of outbreaks, affecting 351 workplaces, as well as the highest rate of outbreaks (155/100,000 workplaces) (Table 2). Of the 151 Upper Tier Local Authorities (UTLAs), the largest numbers of workplace outbreaks were mainly observed in northern English towns and cities with the highest outbreak rates registered in Blackburn with Darwen (387/100,000), Sandwell (351/100,000), Liverpool (349/100,000), Rochdale (277/100,000), Manchester (275/100,000) and Bradford (254/100,000).

Table 2 Number and rate of COVID-19 workplace outbreaks by English Region, May-Oct 2020

Region	Number of Outbreaks	Number of Workplaces	Outbreak Rate[(a)] (per 100,000)
North West	351	226,576	155
Yorkshire and The Humber	198	168,184	118
West Midlands	215	183,534	117
East Midlands	134	156,900	85
North East	39	67,056	58
London	149	375,249	40
South West	84	215,640	39
East of England	71	226,190	31
South East	64	349,945	18
Total	1,305	1,969,274	66

(a): Due to the uncertainties in the data gathered for this analysis, confidence intervals are not presented since this would only represent statistical uncertainty.

Outbreak rates by sector

In comparison with other sectors, retailers had the highest number of outbreaks, affecting 219 workplaces, followed by manufacturers and packers of non-food products (195) and offices (193). However, after applying the denominator data, the highest outbreak rate was in manufacturers and packers of food (1,672/100,000), based on 117 outbreaks out of 6,998 workplaces. This was much higher than the outbreak rates for the remaining sectors with warehouses and manufacturers and packers of non-food products the next highest at 385 per 100,000 workplaces and 308 per 100,000 workplaces respectively (Table 3).

Table 3 Number and rate of workplace outbreaks by sector in England, May-Oct 2020

Sector	Number of Outbreaks	Number of Workplaces	Outbreak Rate (per 100,000)
Manufacturers and packers of food	117	6,998	1,672
Warehouses	58	15,058	385
Manufacturers and packers of non-food	195	63,312	308

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Retailers	219	195,025	112
First responders/Military sites	57	67,257	85
Distributors and transporters	84	125,414	67
Restaurants and caterers	53	117,836	45
Offices	193	721,351	27
Close contact services	13	52,866	25
No setting type assigned	54	511,071	11
Primary producers	8	93,086	9
Other	266	-	-
Total	1,317	1,969,274	67

Outbreak rates by region and sector

High outbreak rates in manufacturers and packers of food were observed consistently across seven regions, including: the West Midlands (3,555/100,000 workplaces), Yorkshire and the Humber (3,132/100,000 workplaces), the North West (2,926/100,000 workplaces), the East Midlands (2,031/100,000 workplace), the East of England (1,664/100,000), the North East (1,282/100,000 workplaces), and the South West (638/100,000 workplaces) (Table 4). In addition, high rates of outbreaks were observed in warehouse settings in the East Midlands and the North West with an outbreak rate of 1,524 per 100,000 workplaces and 793 per 100,000 workplaces respectively (Table 4). See online supplemental Table S1 for more information on the outbreak rate for each combination of region and sector.

Table 4 The top 10 outbreak rates by English region and sector combined, May-Oct 2020

Region - Sector	Number of Outbreaks	Number of Workplaces	Outbreak Rate (per 100,000)
West Midlands - Manufacturers and packers of food	23	647	3,555
Yorkshire and The Humber - Manufacturers and packers of food	28	894	3,132
North West - Manufacturers and packers of food	28	957	2,926
East Midlands - Manufacturers and packers of food	13	640	2,031
East of England - Manufacturers and packers of food	12	721	1,664
East Midlands - Warehouses	19	1,247	1,524
North East - Manufacturers and packers of food	4	312	1,282
North West - Manufacturers and packers of non-food	65	8,074	805
North West - Warehouses	15	1,891	793
South West -	6	940	638

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Manufacturer and packers of food			
…			

Attack rates by enterprise size

A minority (29%) of the outbreaks recorded in SOI were in small enterprises (<50 employees) but the proportion of small enterprises was higher for close contact services (83%) and restaurants and caterers (56%). The overall median attack rate was 3.4% for outbreaks in all enterprises. The median attack rate was 1.1% for outbreaks in large enterprises (250 employees or more), 4.3% in medium size enterprises (50-249 employees), and 17.8% in small enterprises (1-49 employees). The attack rates increased as the number employed at a workplace decreased.

Attack rates by sector

Outbreaks in close contact services had the largest attack rate (median 16.5%), which was based on 22 test-confirmed cases at 6 outbreak sites (Table 5). The attack rates were also high for outbreaks in restaurants and caterers (median 10.3%), based on 49 test-confirmed cases at 14 sites; and in manufacturers and packers of non-food products (median 6.7%), which was based on 270 cases at 29 sites. Most of the outbreaks (162 of 264 outbreaks) had an attack rate less than 6%. However, in a small number of outbreaks (57 of 264) the attack rate was over 15% (see online supplemental figure S2).

Table 5 Median attack rates of workplace outbreaks by sector in England, May-Oct 2020

	Individual Outbreaks				Workplaces		Attack Rate	
Sector	Total cases	Total sites	Cases per site		Number employed at outbreak sites		Cases per 100 employed	
			Median	IQR[(a)]	Median	IQR[(a)]	Median	IQR[(a)]
Close contact services	22	6	3	2	16	10	16.5	6.7
Restaurants and caterers	49	14	4	2	38	94	10.3	14.2
Manufacturers and packers of non-food	270	29	8	7	122	269	6.7	11.1
No setting type assigned	99	15	4	6	56	112	5.4	9.9
Retailers	115	28	4	2	120	242	4.9	16.1
Offices	133	23	5	4	133	207	4.3	15.7
Manufacturers and packers of food	1,384	79	7	12	423	641	2.3	7.0
First Responders/ Military sites	44	15	3	2	113	422	2.1	4.0
Other	109	24	3	3	169	241	2.0	10.6
Warehouses	104	12	3	8	579	781	1.6	1.2
Distributors and transporters	193	16	4	9	650	693	1.2	4.7
Primary producers	127	3	3	61	*	*	*	*
Total	2,649	264	4	6	176	473	3.4	11.3

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(a): Interquartile range (IQR)
*: the number of outbreak sites is too small to calculate.

DISCUSSION

Our study has utilised the number of confirmed COVID-19 outbreaks recorded in PHE information system and combined them with relevant denominator data held by HSE to calculate outbreak rates and attack rates by sector and geographical area. A relatively large number of outbreaks were observed in some workplace settings, including retail, manufacturers and packers of non-food products and offices. After applying the denominator data of the total number of the relevant settings, manufacturers and packers of food had the highest outbreak rates and this was consistent across seven English regions. Manufacturers and packers of food are part of the national infrastructure and these workplaces were kept open throughout the pandemic even during the national lockdown. Outbreaks of COVID-19 in manufacturers and packers of food have been frequently reported in the literature and in the media in many counties[16]. However, only a few studies have investigated the potential transmission risk factors in this type of workplace settings [17]. High rates of outbreaks were found in sectors where production demands are high, and workers cannot work from home. It will be important to continue to monitor outbreak rates by industrial sector as the country is moving out of the pandemic and more sectors are increasing their work capacity.

Our study has also utilised data from the public health COVID-19 outbreak management records to calculate infection attack rates. This allows comparison of the potential extent of transmission between outbreaks in different workplace settings. Close contact services and restaurants/caterers had the highest attack rates which were mostly associated with outbreaks in small enterprises. Manufacturers and packers of non-food products also had relatively large attack rates but were mostly associated with outbreaks in medium and large enterprises. However, it is worth noting that the SOI data are skewed towards large and more impactful outbreaks. Furthermore, more detailed analysis of attack rates is limited by low numbers of outbreaks in certain industrial sectors, such as primary producers which include fruit and vegetable growers, animal and animal products.

Our analysis carried some limitations. The potential under-identification of outbreaks in small enterprises (<50 employees) in the numerator coupled with the vast number of small enterprises in the denominator may greatly underestimate the outbreak rates. This could particularly impact on small business-dominated sectors, such as close contact services and restaurants/caterers, where estimated outbreak rates were relatively low, but attack rates were relatively high.

The number of outbreaks reported to HPZone could be affected by national and local level operational changes. For example, as caseload increased in September and October 2020, some HPTs transferred the management of some outbreaks/clusters to local authorities. As a result, HPZone no longer represents a comprehensive list of COVID-19 outbreaks/clusters in England. This will affect the ability to measure the changes of outbreak occurrence or outbreak rate over time, as well as the ability to measure regional variations but it remains valuable to conduct sector comparisons.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

SOI outbreak data is a subset of the HPZone outbreaks/clusters. Data entry was through a separate mechanism. The proportion of HPZone outbreaks/clusters in the workplace being reported as SOI decreased over time, especially from September 2020 onward as HPTs were under pressure to respond to an increasing number of outbreaks. However, it is unclear if these decreased are biased toward certain sectors. It remains valuable to assess the attack rates of individual outbreaks across different sectors.

NPD workplaces information also has some limitations in providing reliable working population data as the denominator, which will cause imprecisions in the attack rate calculation. In our study, NPD data represent the distribution of the GB population pre-pandemic; the number of employees in some workplace settings will be reduced during the pandemic due to social distancing measures. This may cause underestimation of the attack rates due to overestimation of the denominator. The level of underestimation varies by sector with some sectors completely closed and others kept operating in full capacity throughout the pandemic. However, the impact of this limitation may attenuate as society gradually opens.

In addition, the NPD workplaces information may not capture the number of employees in the transient workforce or working in irregular patterns, for example, seasonal workers in the agriculture sector. Employees in some other workplaces, such as in distribution centres, transportation of goods between depots, and in construction will be accounted for but their non-fixed working locations will not be well-represented by a single geographical reference (e.g. postcode of the company address). Similarly, agency staff and sub-contractors are unlikely to be accounted for at the location where they carry out their work activities. This may cause over estimation of the attack rate due to the underestimation of the denominator.

Early identification of COVID-19 outbreaks/clusters and visualisation of their geographical distribution can provide a rapid assessment of where the SARS-CoV-2 transmission is occurring. A large number of COVID-19 outbreaks/clusters have been reported, both in scientific literature and in the media, in a wide range of mostly indoor settings across the world [3 18]. Most of the COVID-19 clusters will be in residential settings, particularly in households, due to the increased risk of transmission caused by close and frequent contact[19]. However, a household cluster will not result in a large outbreak without the virus spreading beyond the household setting. Some of these individuals in households could also travel to other settings including the workplaces. Transmission is a continuous risk. It is difficult to establish where transmission really occurred. Community transmission will also occur through social gathering, particularly gathering outdoors, shopping in supermarkets or using public transport. However, it is difficult to identify outbreaks/clusters from the large number of transient populations in these settings without a rigorous surveillance system for widespread testing and detailed contract tracing. This may underestimate the relative importance of the potential transmission in these less well-defined settings or population.

Since our study, the approach of utilising the suitable denominator data to calculate outbreak rates has been adopted by the UK Joint Bio-security Centre (JBC) and will be embedded in their regular national surveillance analysis and reporting on workplace outbreaks and outbreaks rates. Although this study was only able to analyse the workplace outbreak data in England, the same approach can be applied to the calculation of outbreak rates and attack rates in other countries in the UK, Europe,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

and United States where the relevant available data sources can be explored. The same approach can also be applied to the calculation for other types of settings, such as care homes, hospitals, schools, and prisons. These will potentially guide interventions to target high risk areas and to limit the spreading of the virus.

This study was not able to assess the potential changes in COVID-19 outbreak rates and attack rates over time due to, in part, the limited time period of data and the inconsistency in recording outbreaks/clusters in the HPZone and SOI datasets. Further consideration will be to analyse the more enhanced outbreak/cluster data collected over time from NHS Test and Trace to identify past and emerging trends.

Evidence shows that there could be marked heterogeneity in the characteristics of SARS-CoV-2 transmission[4], with the majority (~80%) of the secondary transmission caused by a very small proportion of SARS-CoV-2 infected persons, and outbreaks of COVID-19 distributed unevenly in certain settings and geographical locations[20]. Our study has found increased rates of outbreak in certain industrial sectors and geographical regions, and a large variation of the size of the attack rates. The variation of the rates may be impacted by the type of work activities, the size of the enterprises, the transmission risk and the intervention strategies to limit the transmission in these sectors. The risk of transmission will also be associated with the behavioural and social factors of the individuals, the environment and the control measures that influence transmission dynamics of the virus in certain settings[3].

The current study has investigated the patterns and rates of COVID-19 outbreaks in England. Further studies, as part of the National Core Study programme, will investigate and identify the characteristics of the outbreak settings that could increase risk of transmission. A comprehensive epidemiological field study has been designed and commissioned to collect data from live COVID-19 outbreaks in workplace settings to better understand the transmission risk factors and transmission routes[8].

Contributors Conceptualization, Y.C., T.A., and F-M.K.; Methodology, T.A., Y.C., and C.F.F.; Validation, Y.C., T.A., C.F.F., and F-M.K.; Formal Analysis, T.A.; Investigation, Y.C., T.A., and C.F.F.; Data Curation, T.A. and C.F.F.; Writing-Original Draft Preparation, Y.C., T.A., and C.F.F.; Writing-Review & Editing, Y.C., T.A., C.F.F., and F-M.K.; Visualization, T.A.; Supervision, Y.C. and F-M.K.; Project Administration, F-M.K. and Y.C.; Funding Acquisition, UK COVID-19 National Core Studies Consortium.

Ethics approval This study constituted public health surveillance rather than research in human beings. Ethical approval from institutional review boards was not required. All information on individual persons had been anonymised.

Funding This work is supported by funding from the PROTECT COVID-19 National Core Study on Transmission and Environment, managed by the Health and Safety Executive on behalf of HM Government.

Competing interests None declared.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Disclaimer The opinions and assertions contained herein are private views of the authors and do not necessarily reflect those of the Health and Safety Executive or Public Health England.

Copyright © Crown copyright (2021), Health and Safety Executive. This is an open access article distributed under the terms of the Open Government Licence v3.0, which permits re-use, distribution, reproduction and adaptation, provided the original work is properly cited.

Patient consent for publication Not required.

Data Availability Statement: The data used to support the findings of the study are included in the references within this paper. No additional data available.

Acknowledgements
We would like to thank Gillian Frost for her advices on the statistical analysis.

REFERENCES

1. Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* 2021;19(3):141-54. doi: 10.1038/s41579-020-00459-7
2. Coronavirus disease 2019 (COVID-19): situation report, 51. Geneva: World Health Organization, 2020.
3. Leclerc Q, Fuller N, Knight L, et al. What settings have been linked to SARS-CoV-2 transmission clusters? [version 2; peer review: 2 approved]. *Wellcome Open Research* 2020;5(83) doi: 10.12688/wellcomeopenres.15889.2
4. Wong NS, Lee SS, Kwan TH, et al. Settings of virus exposure and their implications in the propagation of transmission networks in a COVID-19 outbreak. *The Lancet Regional Health - Western Pacific* 2020;4:100052. doi: https://doi.org/10.1016/j.lanwpc.2020.100052
5. PHE. National COVID-19 surveillance report: 2 October 2020 (week 40): Pubic Health England (https://www.gov.uk/government/publications/national-covid-19-surveillance-reports), 2020.
6. ECDC. COVID-19 clusters and outbreaks in occupational settings in the EU/EEA and the UK. Stockholm: European Centre for Disease Prevention and Control (ECDC), 2020.
7. Curran A. COVID-19 Outbreak investigation to Understand Transmission (COVID-OUT). 2020. https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/the-covid-out-study-covid-19/.
8. Theme 1: Rapid investigation of outbreaks and evidence synthesis; COVID-19 National Core Study on Transmission and Environment 2021 [Available from: https://sites.manchester.ac.uk/covid19-national-project/rapid-investigation-of-outbreaks-and-evidence-synthesis/.
9. PHE. Weekly national Influenza and COVID-19 surveillance report, Week 6 report (up to week 5 data) in 2021: Public Health England (https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports), 2021.
10. PHE. COVID-19: epidemiological definitions of outbreaks and clusters in particular settings, Public Health England: Pubic Health England; [updated 7 August 2020. Available from: https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters/covid-19-epidemiological-definitions-of-outbreaks-and-clusters-in-particular-settings#fnref:1.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

11. Kara-Zaïtri C, Gelletlie R, Schweiger M. The development and deployment of a national web-based system for communicable disease control in England. *International Journal of Infectious Diseases* 2012;16:e133. doi: https://doi.org/10.1016/j.ijid.2012.05.301
12. Gorce J-P, Aldridge, T., Forder, K., Holmes, W., McManus. H., and Purdy, K. The National Population Database: overview and developments. Health and Safety Executive Research Report RR1130, 2018.
13. Inter-Departmental Business Register (IDBR) Office for National Statistics (ONS) (https://www.ons.gov.uk/aboutus/whatwedo/paidservices/interdepartmentalbusinessregisteridbr), 2021.
14. CDC. Principles of Epidemiology in Public Health Practice: Centers for Disease Control and Prevention, https://www.cdc.gov/csels/dsepd/ss1978/glossary.html, 2012.
15. ONS. ONS Postcode Directory (November 2020). Open Geography Portal: Office for National Statistics (https://geoportal.statistics.gov.uk/datasets/ons-postcode-directory-november-2020), 2020.
16. Waltenburg MA, Rose CE, Victoroff T, et al. Coronavirus Disease among Workers in Food Processing, Food Manufacturing, and Agriculture Workplaces. *Emerg Infect Dis* 2021;27(1):243-49. doi: 10.3201/eid2701.203821
17. Guenther TaC-S, Manja and Indenbirken, Daniela and Robitailles, Alexis and Tenhaken, Peter and Exner, Martin and Ottinger, Matthias and Fischer, Nicole and Grundhoff, Adam and Brinkmann, Melanie,. Investigation of a superspreading event preceding the largest meat processing plant-related SARS-Coronavirus 2 outbreak in Germany *SSRN* 2020 doi: http://dx.doi.org/10.2139/ssrn.3654517
18. Liu T, Gong D, Xiao J, et al. Cluster infections play important roles in the rapid evolution of COVID-19 transmission: A systematic review. *International Journal of Infectious Diseases* 2020;99:374-80. doi: https://doi.org/10.1016/j.ijid.2020.07.073
19. PHE. National flu and COVID-19 surveillance reports, Public Health England: Public Health England; [updated 25 March 2021. Available from: https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports.
20. Frieden T, Lee C. Identifying and Interrupting Superspreading Events—Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Disease journal* 2020;26(6):1059. doi: 10.3201/eid2606.200495

For Peer Review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure S1 Geographical distributions of COVID-19 outbreaks associated with the workplace in England, May – Oct 2020

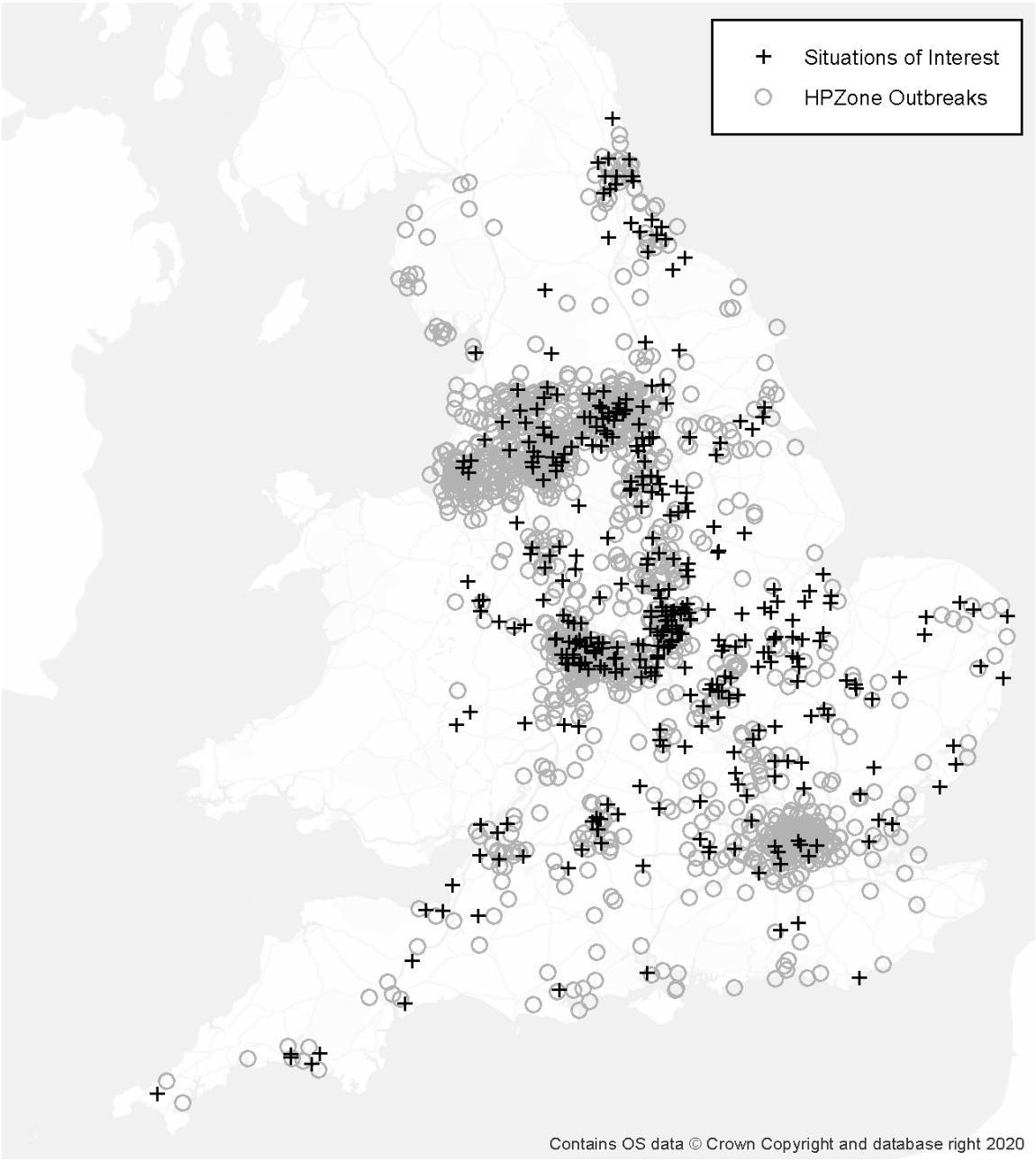


Figure S2 Distribution of the number of workplace outbreaks by attack rate in England, May-Oct 2020

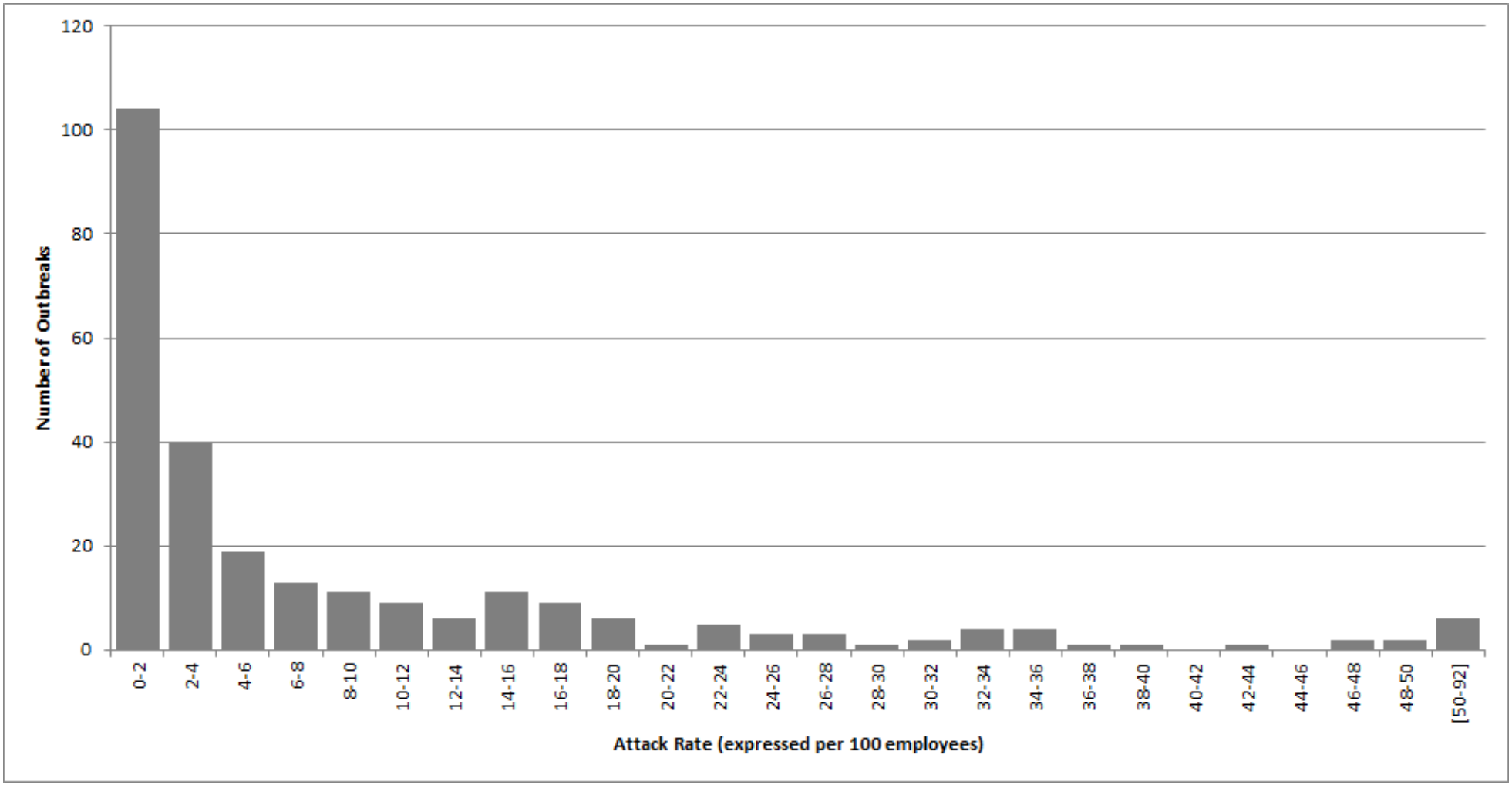


Table S1 Number and rate of COVID-19 workplace outbreaks by English Region and Sector, May-Oct 2020
The outbreak rates in italics indicate where they were calculated from numbers of outbreaks that were less than 10.

		Region										
Sector	**Statistic**	**East Midlands**	**East of England**	**London**	**North East**	**North West**	**South East**	**South West**	**West Midlands**	**Yorkshire and The Humber**	**No Region Assigned**	**Total**
Close contact services	Outbreak Rate (per 100,000)	*0*	*0*	*22*	*0*	*29*	*31*	*40*	*0*	*89*		25
	Number of Outbreaks	*0*	*0*	*2*	*0*	*2*	*3*	*2*	*0*	*4*		13
	Number of Workplaces	4228	5959	9163	2204	6821	9676	5026	5279	4510		52866
Primary producers	Outbreak Rate (per 100,000)	*0*	*35*	*0*	*0*	*9*	*0*	*0*	*25*	*0*		*9*
	Number of Outbreaks	*0*	*4*	*0*	*0*	*1*	*0*	*0*	*3*	*0*		*8*
	Number of Workplaces	10639	11323	434	3669	10979	10564	22434	11966	11078		93086
Distributors and transporters	Outbreak Rate (per 100,000)	*33*	*26*	*40*	*54*	117	*35*	*51*	146	112		67
	Number of Outbreaks	*4*	*4*	*8*	*2*	18	*7*	*6*	21	14		84
	Number of Workplaces	12042	15573	20046	3731	15336	20148	11664	14403	12471		125414
First responders/ Military sites	Outbreak Rate (per 100,000)	*84*	*13*	92	*0*	116	*35*	*125*	*92*	188		85
	Number of Outbreaks	*5*	*1*	10	*0*	10	*4*	*9*	*6*	12		57
	Number of Workplaces	5938	7596	10831	2860	8640	11271	7217	6515	6389		67257
Manufacturers and packers of food	Outbreak Rate (per 100,000)	2031	1664	*204*	*1282*	2926	*110*	*638*	3555	3132		1672
	Number of Outbreaks	13	12	*2*	*4*	28	*1*	*6*	23	28		117
	Number of Workplaces	640	721	982	312	957	905	940	647	894		6998
Manufacturers and packers of non-food	Outbreak Rate (per 100,000)	283	*78*	*102*	464	805	*79*	215	461	406		308
	Number of Outbreaks	19	*6*	*7*	11	65	*8*	14	38	27		195
	Number of Workplaces	6714	7710	6860	2372	8074	10163	6520	8246	6653		63312
Offices	Outbreak Rate (per 100,000)	30	*10*	19	58	81	*6*	*13*	41	46		27
	Number of Outbreaks	14	*8*	35	11	62	*8*	*9*	23	22	*1*	193
	Number of Workplaces	47025	78736	188040	19015	76250	140530	66957	56551	48247		721351
Restaurants and caterers	Outbreak Rate (per 100,000)	*45*	*33*	64	*19*	99	*0*	*7*	*56*	*71*		45
	Number of Outbreaks	*4*	*4*	14	*1*	14	*0*	*1*	*6*	*8*	*1*	53

For peer review only

	Number of Workplaces	8893	11944	21979	5291	14145	19211	14364	10680	11329		117836
Retailers	Outbreak Rate (per 100,000)	166	49	59	*36*	225	*22*	56	216	205		112
	Number of Outbreaks	26	10	20	*3*	57	*7*	12	42	39	*3*	219
	Number of Workplaces	15635	20537	33739	8269	25339	31753	21260	19457	19036		195025
Warehouses	Outbreak Rate (per 100,000)	1524	*99*	*40*	*0*	793	*0*	*379*	*629*	*399*		385
	Number of Outbreaks	19	*2*	*1*	*0*	15	*0*	*5*	*9*	*6*	*1*	58
	Number of Workplaces	1247	2018	2475	500	1891	2675	1319	1431	1502		15058
Not assigned	Outbreak Rate (per 100,000)	*16*	*8*	*6*		21	*4*	*2*	*14*	*17*		11
	Number of Outbreaks	*7*	*5*	*5*		12	*4*	*1*	*7*	*8*	*5*	54
	Number of Workplaces	43899	64073	80700	18833	58144	93049	57939	48359	46075		511071
Other	Outbreak Rate (per 100,000)	-	-	-	-	-	-	-	-	-		
	Number of Outbreaks	23	15	45	*7*	67	22	19	37	30	*1*	266
	Number of Workplaces	0	0	0	0	0	0	0	0	0		
Total	Outbreak Rate (per 100,000)	85	31	40	58	155	18	39	117	118		**67**
	Number of Outbreaks	134	71	149	39	351	64	84	215	198	12	**1317**
	Number of Workplaces	156900	226190	375249	67056	226576	349945	215640	183534	168184		**1969274**

For peer review only

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 1-2	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Page 2
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 2-3		
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 3		
Methods					
Study Design	4	Present key elements of study design early in the paper	Page 3		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 3		

For peer review only

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	Page 3-5	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Page 3-5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Page 4-5	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Page 5
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Page 3-4		

For peer review only

Bias	9	Describe any efforts to address potential sources of bias	Page 5-6		
Study size	10	Explain how the study size was arrived at	Page 5-6		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Page 4-6		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Page 4-6		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	Page 3-5

For Peer review only

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Page 5-6
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Page 6	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Page 6
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	Page 6		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure	Page 6		

For peer review only

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Page 6-9		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			
Discussion					
Key results	18	Summarise key results with reference to study objectives	Page 9-10		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 10-11	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Page 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives,	Page 9-12		

For peer review only

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 11-12		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 12		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 12

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

*Checklist is protected under Creative Commons Attribution (CC BY) license.

For peer review only